Journal of Biomedical Research

The single-cell atlas of programmed cell death signature: A machine learning-based prognostic framework in breast cancer

Jixuan Jiang, Yiwen Wang, Yuju Huang, Weiwen Yan, Shuwei Li, Ruoxi Wang

Cite this article as:
Jixuan Jiang, Yiwen Wang, Yuju Huang, Weiwen Yan, Shuwei Li, Ruoxi Wang. The single-cell atlas of programmed cell death signature: A machine learning-based prognostic framework in breast cancer[J]. *Journal of Biomedical Research*, In press. doi: 10.7555/JBR.40.20260038

View online: https://doi.org/10.7555/JBR.40.20260038

Available online at www.jbr-pub.org.cn

Open Access at PubMed Central

Journal of Biomedical Research, 2026 40(0): 1–21

JBR

Original Article

The single-cell atlas of programmed cell death signature: A machine learning-based prognostic framework in breast cancer

Jixuan Jiang[1,△], Yiwen Wang[2,△], Yuju Huang[1,△], Weiwen Yan[2], Shuwei Li[2,✉], Ruoxi Wang[1,✉]

[1]Department of Breast Surgery, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu 210000, China;
[2]Department of Environmental Genomics and Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China.

Abstract

Breast cancer remains a leading cause of cancer-related mortality in women, and current prognostic models are suboptimal. The transcriptomic role of programmed cell death (PCD) in breast cancer progression is not fully understood. Here, we integrated single-cell RNA sequencing data from breast tumors with nine bulk transcriptomic cohorts to systematically analyze 19 PCD modalities. Using a machine learning framework incorporating 14 algorithms, we constructed a prognostic signature, with a ridge regression-based PCD riskscore showing optimal performance and being further integrated into a clinical nomogram. Functional roles of key genes were validated through *in vitro* and *in vivo* experiments. We identified a prognostic signature comprising 26 core PCD genes, which effectively stratified patients into distinct risk groups and robustly predicted overall survival. Single-cell analyses revealed that a high PCD risk core was associated with an immunosuppressive tumor microenvironment and reduced immune checkpoint expression, whereas low-risk patients showed greater sensitivity to targeted therapies. Among the signature genes, *PDIA4* was consistently overexpressed in 50 paired breast cancer tissues, and its knockdown markedly inhibited tumor growth and malignant phenotypes. This study establishes a novel PCD-based prognostic signature for breast cancer and identifies *PDIA4* as a functionally important oncogene.

Keywords: programmed cell death, machine learning, prognostic signature, breast cancer, *PDIA4*

Introduction

Breast cancer is the predominant malignancy among females globally, with over 2.3 million new diagnoses in 2020[1]. It ranks as the second most prevalent cause of cancer-related mortality in females, behind lung cancer[2], due to considerable diagnostic delay and failure of therapeutic strategies. Currently, the identification of valid biomarkers, particularly new molecular therapeutic targets, has been the focal point

△These authors contributed equally.

✉Corresponding authors: Shuwei Li, Email: shuweili@njmu.edu.cn, ORCID: 0000-0002-6142-0573; Ruoxi Wang, Email: strawberry210705@163.com, ORCID: 0000-0002-7865-4405.

Received: 16 January 2026; Revised: 18 March 2026; Accepted: 01 April 2026; Available online: 04 July 2026

CLC number: R737.9, Document code: A

The authors reported no conflict of interests.

https://doi.org/10.7555/JBR.40.20260038

of breast cancer studies[3]. Previous studies have used machine learning models to explore predictive biomarkers of breast cancer[4]. However, tumor heterogeneity leads to significant differences in patient response to immunotherapy and targeted therapy[5]. Current common parameters guiding treatment decisions, such as tumor size and grade, nodal involvement, histopathology, and marker expression, remain suboptimal[6]. Therefore, it is imperative to develop prognostic models to predict the effectiveness of targeted and immune therapies in breast cancer .

Cell death is divided into accidental cell death and programmed cell death (PCD). PCD denotes the self-governed, sequential demise of cells directed by genes and regulated by the creation of signaling amplification complexes[7]. PCD involves diverse mechanisms of cellular death, such as autophagy, apoptosis, alkaliptosis, cuproptosis, disulfidptosis, entotic cell death, ferroptosis, lysosome-dependent cell death, necroptosis, NETotic cell death, oxeiptosis, pyroptosis, and parthanatos[8-10].

Tumor cells exploit PCD dysregulation to promote survival, metastasis, and therapy resistance[11]. Recent pan-cancer studies demonstrate that PCD signatures predict prognosis and drug sensitivity in lung[12] and renal carcinomas. Pyroptosis can inhibit tumor development by eliminating aberrant cells in the early stages of cancer, but promote tumor progression by promoting inflammation and tissue injury in advanced cancer[13]. *DUSP16* has been reported to reduce the accumulation of *BAX* in mitochondria, thereby inhibiting apoptosis to promote chemoresistance[14]. Notably, increasing evidence suggests that autophagy modulation may serve as a viable strategy for anti-triple-negative breast cancer (TNBC) drug development[15]. Autophagy regulation is associated with numerous signaling pathways and target proteins, including the unc-51-like kinase 1 complex (*ULK1*), *PI3K-AKT-mTORC1*[16], fork-head box O (*FOXO*), Ras-Raf-MAPKs, p53, Beclin-1, Plat-Erk1/2, p62, and NF-κB[17]. Nevertheless, few investigations have comprehensively elucidated the association between PCD patterns and breast cancer prognosis, immune infiltration, and treatment efficacy. Although studies have revealed the significant role of PCD in postoperative prognosis and drug sensitivity in TNBC[18], there remains a lack of validation through *in vitro* and *in vivo* experiments. Therefore, we aim to integrate multi-omics data to investigate the key functions of PCD within the tumor microenvironment. Integrating multi-omics data offers unprecedented resolution to reveal the key roles of PCD in tumor microenvironment remodeling and immune evasion.

In this study, we detected 26 differentially expressed genes related to PCD in breast cancer *via* bulk transcriptomics and single-cell RNA-sequencing (scRNA-seq) datasets and used machine learning to develop a PCD risk score associated with breast cancer prognosis through an in-depth analysis of large-scale data. Additionally, we investigated the relationship between the PCD risk score and immune infiltration and targeted therapy in breast cancer. Both immunohistochemical analysis and malignant phenotype assays were conducted to verify the expression and function of key PCD genes (***Fig. 1***).

Materials and methods

Data acquisition

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, gene sets from Gene Set Enrichment Analysis (GSEA), and pertinent review articles were compiled to represent 3 346 genes linked to 19 different patterns of PCD. These 19 modalities were identified based on standardized classifications from the Nomenclature Committee on Cell Death (NCCD) and well-established literature. To evaluate the redundancy of the gene sets, a Jaccard similarity analysis was conducted to quantify the gene-level overlap among these modalities. As training datasets, bulk RNA-seq data from breast cancer tissue samples and associated clinicopathologic information were sourced from the Gene Expression Omnibus (GEO; GSE88770, GSE42568, and GSE58644) and The Cancer Genome Atlas (TCGA). Five datasets were used as validation datasets, including GSE103091, GSE20711, GSE21653, GSE24450, and METABRIC. The GSE161529 cohort, which is accessible through the GEO database, provided the scRNA-seq data. Transcripts Per Million (TPM)-formatted data were used for model training and validation. Missing values were omitted. Overall survival was identified as the primary endpoint across all datasets. To integrate multiple cohorts, batch effects were corrected using the "ComBat" algorithm in the "sva" package. The expression data were $\log_2$-transformed [$\log_2$(TPM + 1)], and the study origin was defined as the batch variable for adjustment (***Supplementary Fig. 1A***). Specifically, a model matrix was constructed to preserve biological variation, and the parametric empirical Bayes method was applied by setting the "par.prior" parameter to TRUE.

Determination of the expression and variation levels of PCD-associated genes

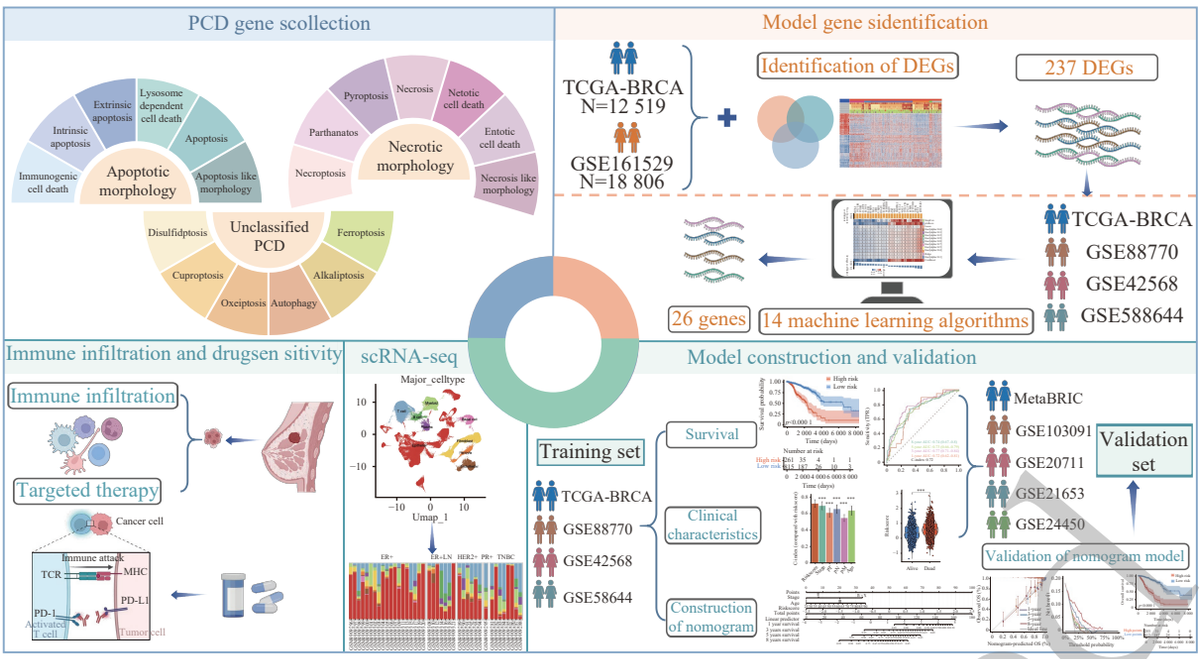


Fig. 1 **Flow chart of the study.** Comprehensive evaluation of programmed cell death and integrated development of breast cancer signatures to predict prognosis and therapy response.

Raw count data downloaded from The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) were preprocessed and analyzed for differential expression using the "DESeq2" package. Differentially expressed genes (DEGs) were defined as $|\log_2(\text{fold change [FC]})|>1$ with an adjusted *P*-value (adj*P*) < 0.05. The *P*-value adjustment was performed using the Benjamini–Hochberg method to control the false discovery rate (FDR). Furthermore, scRNA-seq data from GSE161529 were preprocessed and analyzed using the "Seurat" package. DEGs were calculated based on an adjusted $P < 0.01$ and a proportion of cells expressing the gene (pct) > 0.15. To identify PCD-related genes, genes that were consistently upregulated or downregulated in both datasets were intersected with genes from PCD modalities. To identify possible pathways associated with these DEGs, enrichment analysis was then carried out. The "MAFTools" package was used to further investigate the somatic mutation landscape in the cohort of patients with breast cancer. After excluding low-expression genes, 237 PCD-associated DEGs were incorporated into the following analysis.

Development of a PCD-related prognostic risk score based on an integrated machine learning approach

We used 14 linear models, including StepCox, plsRcox, LASSO, Ridge, CoxBoost, and Elastic Net (ENET), to examine four training datasets (TCGA-BRCA, GSE88770, GSE42568, and GSE58644) to determine PCD-associated genes that are significant predictors of survival. For a gene to be considered a significant gene (model gene), the following criteria had to be satisfied simultaneously: (a) the absolute value of its coefficient had to be greater than 0.01 in at least 10 of the 14 models; and (b) the mean of the absolute coefficient values across all models had to be greater than 0.05.

We established breast cancer prognostic models using these genes and obtained a consensus set of model genes with excellent accuracy and stability using the 14 linear algorithms. (1) Using the 26 model genes, predictive models were constructed in the METABRIC cohort *via* leave-one-out cross-validation (LOOCV); (2) these models underwent further cross-validation with four separate datasets (GSE24450, GSE103091, GSE21653, and GSE20711); and (3) Harrell's concordance index (C-index) was calculated for each model across validation datasets. The highest mean C-index determined the most accurate model.

Functional enrichment analysis

The R software package "clusterProfiler" was employed for functional enrichment analysis to identify relevant Gene Ontology (GO) terms using the model genes mentioned above. Biological functions of the high-risk and low-risk groups were contrasted using Gene Set Variation Analysis (GSVA) and the "h.all.v2022.1.Hs.symbols.gmt" gene set collection.

The activity of PCD-related gene sets across individual samples was quantified using single-sample Gene Set Enrichment Analysis (ssGSEA).

Validation of predictive models

Using the "survminer" package and the "pROC" package, four training datasets (TCGA-BRCA, GSE58644, GSE42568, GSE88770) were subjected to Kaplan–Meier analysis and receiver operating characteristic (ROC) curve analysis. Subsequently, we performed the same analysis on five validation datasets (METABRIC, GSE103091, GSE20711, GSE21653, and GSE24450) to assess the prognostic performance of the model. Prediction Analysis of Microarray 50 (PAM50) intrinsic subtypes and risk of recurrence (ROR-S) scores were determined using the "genefu" R package. Subtype assignment and score calculation were performed based on a predefined centroid model. To evaluate prognostic superiority, the C-index of the PCD risk score was benchmarked against the established PAM50 signature across the cohorts.

Nomogram construction and evaluation relying on the PCD risk score

To validate the risk score as a standalone prognostic indicator for breast cancer, we conducted both univariate and multivariate Cox regression analyses to evaluate its association with pertinent clinical characteristics. Then, using the "rms" and "regplot" R packages, we constructed prognostic nomograms from the TCGA-BRCA and GEO databases. Subsequently, we evaluated the performance of the nomograms using calibration curves, decision curve analysis (DCA), and ROC curves. Additionally, the prognostic predictive ability of the nomograms was tested using sample data from three datasets (TCGA-BRCA, GSE58644, and METABRIC). Baseline information for these datasets is shown in ***Supplementary Table 1***.

ScRNA-seq data analysis of prognostic model genes

ScRNA-seq data for 38 breast cancer samples and 13 normal samples were obtained from the GEO database (GSE161529). We processed the scRNA-seq data using the standard workflow of the "Seurat" R package (v4.3.0). Quality control was performed by filtering cells based on the following widely adopted criteria: cells with unique gene counts (nFeature_RNA) between 200 and 6 000, and mitochondrial gene ratio (percent.mt) below 15%. These thresholds effectively remove low-quality cells and potential doublets while preserving biological signals. Subsequently, data normalization, scaling, dimensionality reduction, and clustering were performed using the standard "FindNeighbors", "FindClusters", and "RunUMAP" functions. We investigated the expression of model genes within single cells. Additionally, we determined the signature-specific score (CDIscore) of each cell type using the "AddModuleScore" function. To further investigate molecular mechanisms, pseudotime trajectory analysis was performed using the "monocle3" R package to determine the temporal sequence of gene expression. Additionally, cell-cell communication analysis was conducted using the "CellChat" R package to identify ligand–receptor interactions driven by *PDIA4*.

Analysis of tumor microenvironment and drug sensitivity

We further acquired and computed the association between the expression of model genes and the degree of immune cell infiltration using algorithms including Tumor Immune Estimation Resource (TIMER2.0)[19], Estimating the Proportion of Immune and Cancer cells (EPIC)[20], Microenvironment Cell Populations-counter (MCP-counter)[21], Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE)[22], and Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT)[23]. Subsequently, based on the Tumor Immune Dysfunction and Rejection (TIDE) algorithm[24] and the "oncoPredict" tool, we investigated the associations between the PCD risk score and immunotherapy and targeted therapy.

Clinical breast cancer tissues

A total of 50 breast cancer tissues and paired adjacent normal tissues were obtained from patients at The First Affiliated Hospital of Nanjing Medical University. All participants provided written informed consent before enrollment. Telephone follow-up was conducted to obtain postoperative survival and progression data. Patients were included based on pathological confirmation of breast cancer, no prior anticancer treatment, and the absence of other malignancies or severe perioperative complications. The research protocol received approval from the Institutional Review Board of The First Affiliated Hospital with Nanjing Medical University (approval No. 2022-SR-117).

RT-qPCR of selected genes

Total RNA was extracted from clinical tissue samples or cells using TRIzol reagent (Cat. #15596026,

Invitrogen, Carlsbad, CA, USA) according to the manufacturer's guidelines. cDNA was synthesized from 1 μg of total RNA per sample using the Applied Biosystems TaqMan reverse transcription kit. qRT-PCR analysis was conducted on an Applied Biosystems real-time PCR system with SYBR Premix Ex Taq (Cat. #RR420A, Takara, Kusatsu, Shiga, Japan).

Western blot

Total proteins were extracted using RIPA buffer supplemented with protease inhibitors (both from Epizyme, Shanghai, China). Following extraction, the proteins were separated by 10% SDS-PAGE and transferred onto polyvinylidene fluoride (PVDF) membranes. The membranes were then probed with a primary antibody against *PDIA4* (1∶1 000; Cat. #14712-1-AP, Proteintech, Rosemont, IL, USA) and a corresponding secondary antibody. Protein bands were finally visualized using ECL reagents (Cat. #SQ201, Epizyme, Shanghai, China).

Immunohistochemical (IHC) staining

Breast tissue sections were formalin-fixed and paraffin-embedded, then treated with xylene to remove paraffin and rehydrated through a series of graded ethanol concentrations.

Endogenous peroxidase was quenched using 3% H_2O_2 (10 min, room temperature). Microwave-mediated antigen retrieval was carried out in citrate buffer (10 mmol/L, pH 6.0). After blocking with 10% goat serum (30 min, room temperature), sections were incubated overnight at 4 ℃ with primary antibodies: anti-*PDIA4* (1∶200; Cat. #14712-1-AP, Proteintech), and anti-GAPDH (1∶500; Cat. #No.60004-1-Ig, Proteintech). Following phosphate-buffered saline (PBS) washes, nuclei were visualized by applying horseradish peroxidase (HRP)-conjugated secondary antibodies for 30 minutes at room temperature, followed by hematoxylin counterstaining. Quantitative expression of staining intensity was performed using the H-score. By multiplying the staining intensity score (ranging from 0–3) by the proportion of positive cells (ranging from 0–4), a composite score ranging from 0 to 12 was obtained, which was designated as the IHC score.

Hematoxylin and eosin (H&E) staining

Paraffin-embedded tumor sections (5-μm thick) were deparaffinized in xylene and rehydrated through a series of graded ethanol. The sections were stained with hematoxylin solution (Cat. #KGA223, KeyGen BioTech, Nanjing, China) for 5 min, rinsed with running tap water, and subsequently stained with eosin solution (Cat. #KGA231, KeyGen BioTech) for 2 min. Following dehydration and clearing, the slides were mounted with neutral resin for histopathological examination.

Cell lines and cell culture

Human breast cell lines were sourced from the American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37 ℃ under 5% CO_2 conditions. No contamination was observed following mycoplasma testing.

Knockdown of *PDIA4* in MCF-7 and MDA-MB-231 cells

To establish stable *PDIA4*-knockdown MCF-7 and MDA-MB-231 cell lines, three small interfering RNAs (siRNAs) targeting the human *PDIA4* gene were first designed and synthesized (***Supplementary Table 2***). To identify the most effective sequence, these siRNAs were individually transfected into cells using Lipofectamine 2000 (Cat. #11668019, Invitrogen), and knockdown efficiency was evaluated by qPCR at 48 hours post-transfection. The most effective siRNA sequence was then used to generate a corresponding short hairpin RNA (shRNA), which was cloned into the pLKO.1-puro lentiviral vector. Stable *PDIA4*-knockdown cell pools were then generated by lentiviral infection followed by puromycin selection, and the knockdown efficiency was confirmed by qPCR and Western blot analyses.

Cell proliferation assays

Cell growth was evaluated using a Cell Counting Kit-8 (CCK-8) assay and a 5-ethynyl-2'-deoxyuridine (EdU) incorporation assay. Cells (5×10^4 cells per well) were seeded in 96-well plates, and 10 μL of CCK-8 solution was added to each well after 24, 48, and 72 h, followed by a 2-hour incubation. Absorbance at 450 nm was measured, and the Countess II (AMQAX1000, Thermo Fisher Scientific, Waltham, MA, USA) was used to calculate cell numbers at various time points. After incubation with 50 mmol/L EdU for 2 h at 37 ℃ in 24-well plates, cells were fixed with 4% paraformaldehyde and permeabilized with 0.1% Triton X-100 in PBS. The EdU-positive cell rate was defined as the ratio of EdU-positive cells to DAPI-positive cells.

Colony formation assay

One hundred microliters of a cell suspension at a concentration of 5×10^5 cells/mL was combined with

1 mL of methylcellulose semisolid medium and plated in 6-well plates. After uniform distribution, plates were cultured at 37 ℃ for 15 days.

Migration and invasion assays

Migration and invasion were evaluated using a 24-well Transwell system (Biofil, Guangzhou, China). The upper chamber was seeded with 1×10^5 cells in 200 μL of serum-free DMEM, while the lower chamber was filled with 500 μL of DMEM containing 20% FBS. Following 24 h at 37 ℃, cells that did not migrate or invade were removed. The cells on the lower membrane surface were fixed with 4% paraformaldehyde for 15 min, stained with 0.25% crystal violet for 30 min, and then imaged.

Cell apoptosis analysis

After staining cells with 5 μL of Annexin V-fluorescein isothiocyanate (FITC) and 5 μL of propidium iodide (PI) for 15 min at room temperature, apoptosis was detected using flow cytometry. All steps followed the manufacturer's protocol (Cat. #KGA108, KeyGenBioTech).

Animal xenograft model and ethical statement

For the xenograft assay, *PDIA4*-knockdown or control MCF-7 cells (1×10^6) suspended in 100 μL of PBS were subcutaneously inoculated into the axillary region of female BALB/c nude mice (4–5 weeks old). Prior to inoculation, mice were anesthetized by intraperitoneal injection of 1% pentobarbital sodium solution at a dose of 40–60 mg/kg body weight. Tumor volume and body weight were monitored every three days. At the experimental endpoint (30 days post-inoculation), mice were euthanized by cervical dislocation under deep anesthesia, and tumors were harvested and weighed. Tumor tissues were fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned for H&E staining as well as immunohistochemical analysis. All animal procedures were performed in compliance with protocols approved by the Institutional Animal Care and Use Committee of Nanjing Medical University (IACUC-2512076).

Statistical analysis

Categorical data were analyzed using the chi-square test. Student's *t* test or Wilcoxon rank sum test was applied to normally and non-normally continuous variables, respectively. All data analyses were performed with R software (version 4.3.1). Statistical significance was set at $P < 0.05$.

Results

Flowchart of this study

The dataset information and study workflow are shown in ***Fig. 1***. We further performed comprehensive analysis and screening of various published cohorts to construct and verify our predictive models. The study incorporated four training sets (TCGA-BRCA, GSE88770, GSE42568, and GSE58644) , five validation sets (METABRIC, GSE103091, GSE20711, GSE21653, and GSE24450), and one single-cell RNA-seq dataset (GSE161529).

Identification of differentially expressed PCD-associated genes

We first used ssGSEA to evaluate the cooperative activity patterns and correlations of gene sets related to PCD in breast cancer. This analysis revealed extensive interrelationships among these gene sets (***Fig. 2A***), suggesting their potential synergistic involvement in tumor biological processes. To ensure non-redundancy, we evaluated the gene-level overlap among these modalities (***Supplementary Fig. 1B***). The mean Jaccard index was 0.003 6, confirming that these PCD patterns are genetically independent despite their functional synergy. This justifies the integration of 19 distinct biological dimensions in our prognostic framework. The 19 PCD modalities collectively comprised 3 346 unique genes after merging overlaps. Examples of modality-specific gene counts included 902 from disulfidptosis, 556 from necroptosis, 520 from autophagy, and 500 each from extrinsic apoptosis, immune cell death, intrinsic apoptosis, and necrosis (***Fig. 2B***). Subsequently, we investigated breast cancer expression profiles at bulk and single-cell RNA-seq resolution. A total of 12 519 statistically significant DEGs (adjusted $P < 0.05$ and $|\log_2 FC| > 1$) were identified from the TCGA bulk RNA-seq database (***Fig. 2C***), and 18 806 genes were obtained from the single-cell dataset GSE161529 using a cutoff of adjusted $P < 0.01$ and a minimum cell fraction of 15%. The cytograms of the nine cell subtypes were shown in ***Fig. 2D–2F***. The volcano plot shows the expression of marker genes for the nine cell subtypes (***Fig. 2G***). We identified 146 genes by overlapping PCD genes with genes that were upregulated in both bulk and single-cell RNA-seq data (***Fig. 2H***). Similarly, we combined genes that were downregulated in both datasets with those associated with PCD, obtaining 94 genes (***Fig. 2I***). In total, 240 differentially expressed PCD genes were incorporated into the subsequent analysis. Detailed results are presented in ***Supplementary Table 3***.

Variant landscape of PCD genes in breast cancer

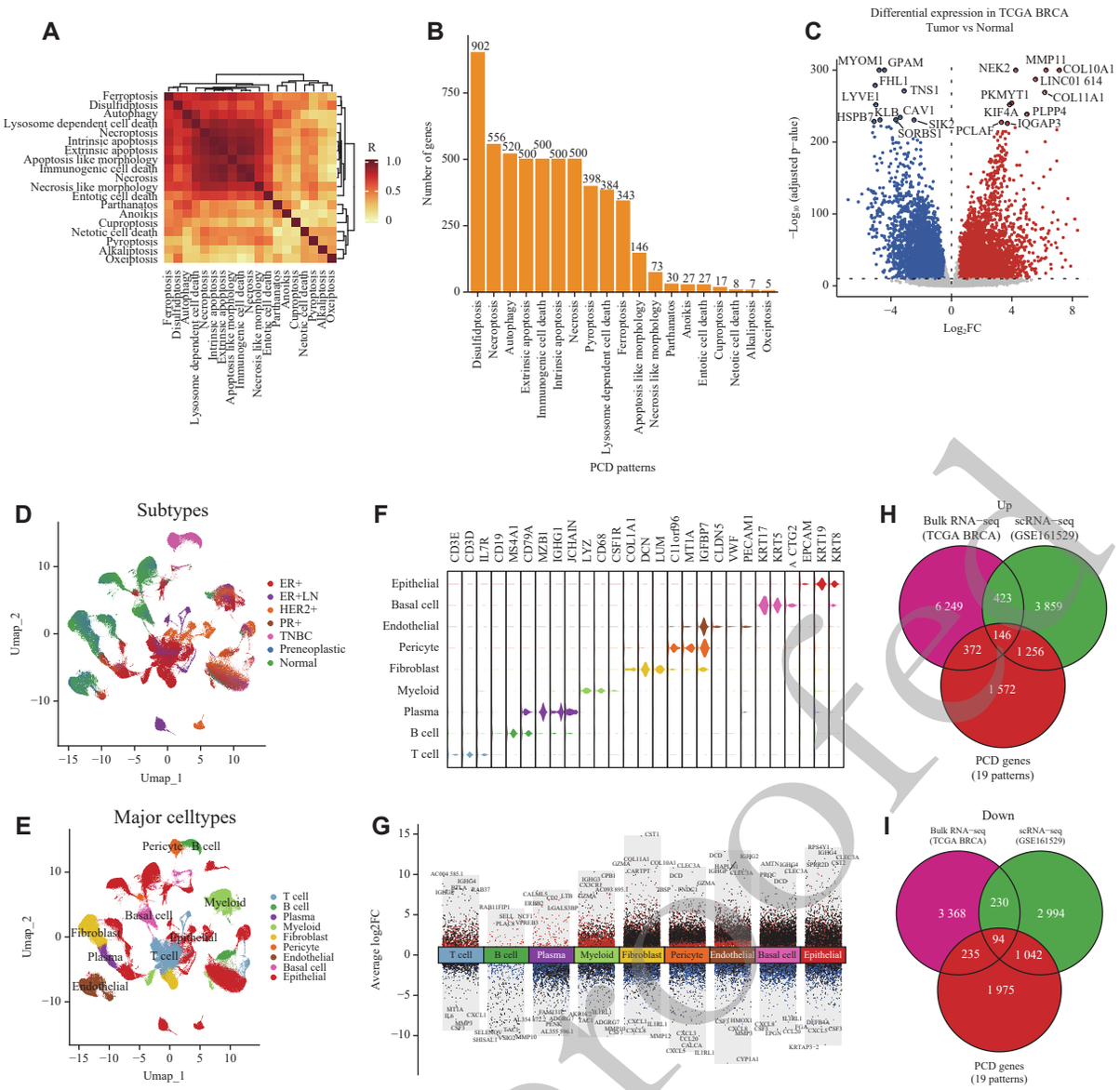


Fig. 2 **Identification of programmed cell death (PCD)-differentially expressed genes (DEGs).** A: Correlation and interaction among 19 PCD patterns. B: Distribution of the number of genes in 19 models. C: Volcano plot showing the DEGs between tumor and normal tissues in TCGA breast cancer samples. D and E: UMAP plots showing the distribution of cell subtypes and major cell types in the single-cell dataset. F: Expression of marker genes in cell subtypes. G: Volcano plot of differentially expressed genes across the nine cell subtypes. H and I: Venn plot of the upregulated (H) and downregulated (I) genes shared among bulk RNA-sequence, scRNA-sequence and PCD genes.

patients

Heatmaps of the scaled RNA levels of DEGs are depicted in ***Fig. 3A***. Additionally, the distribution of the 240 DEGs across the 19 PCD patterns is depicted in ***Fig. 3B***. Hallmark and KEGG enrichment analyses indicated that these DEGs participate in different carcinogenesis-related pathways, such as the TNFα-NF-κB signaling pathway, epithelial-mesenchymal transition, and pathways in cancer (***Fig. 3C–3D***). Furthermore, we explored the mutational landscape of PCD-DEGs within the TCGA breast cancer dataset. The analysis focused on the first 10 mutations in PCD-associated genes. *ERBB2* and *PIK3R1* stood out with the highest mutation frequencies (7%). The other eight genes exhibited mutation frequencies that were relatively lower, falling between 3% and 5% (***Fig. 3E–3F***).

Identification of model genes and construction of a prognostic model through a machine learning-based integrative procedure for breast cancer patients

After eliminating three low-expression genes, 237 differentially expressed PCD-related genes were included in further analysis. We utilized 14 linear model algorithms to identify significant genes based

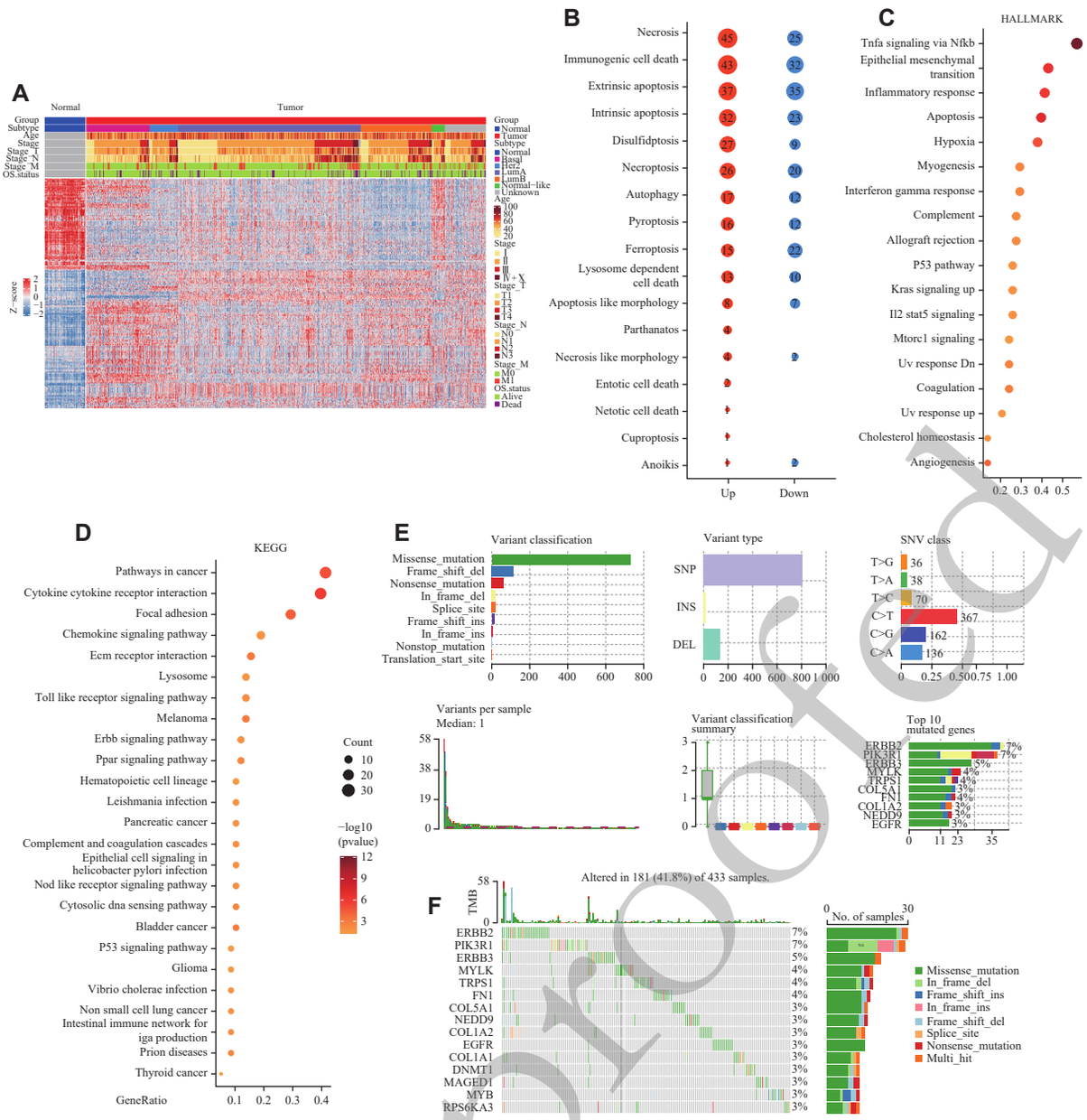


Fig. 3 **Landscape of programmed cell death (PCD)-differentially expressed genes (DEGs) in breast cancer patients.** A: Heatmap of 240 PCD-DEGs and clinical factors. B: Upregulated and downregulated genes contained in 19 PCD pathways. C: HALLMARK enrichment analyses based on the PCD-DEGs. D: KEGG enrichment analyses based on the PCD-DEGs. E and F: Diagram of somatic mutations for the PCD-DEGs.

on expression profiles from four training datasets (TCGA-BRCA, GSE88770, GSE42568 and GSE58644). A total of 26 genes were identified as significant prognostic markers (***Fig. 4A***). Detailed results are presented in ***Supplementary Fig. 2A***. These 26 genes were primarily associated with immunogenic cell death, intrinsic and extrinsic apoptosis, necrosis, and pyroptosis (***Fig. 4B***). The chromosomal location of each gene is depicted in ***Fig. 4C***, and a correlation heatmap of these genes is shown in ***Supplementary Fig. 2B***.

Subsequently, we developed a prognostic risk score using machine learning techniques, leveraging the expression profiles of the 26 model genes. Furthermore, we applied the LOOCV framework to construct 14 prediction models and calculated the C-index for each model across the validation datasets (***Supplementary Table 4***). Notably, the Ridge regression model achieved the top C-index (***Fig. 4D***). Using Ridge regression, the risk score for each patient was calculated from the $\log_2$(TPM+1)-normalized expression values (exp) as follows:

$$\text{Riskscore} = (0.154 \times EIF4EBP1^{exp}) + (0.261 \times EZR^{exp}) + (0.199 \times MEG3^{exp}) + (0.125 \times B4GALT3^{exp}) +$$

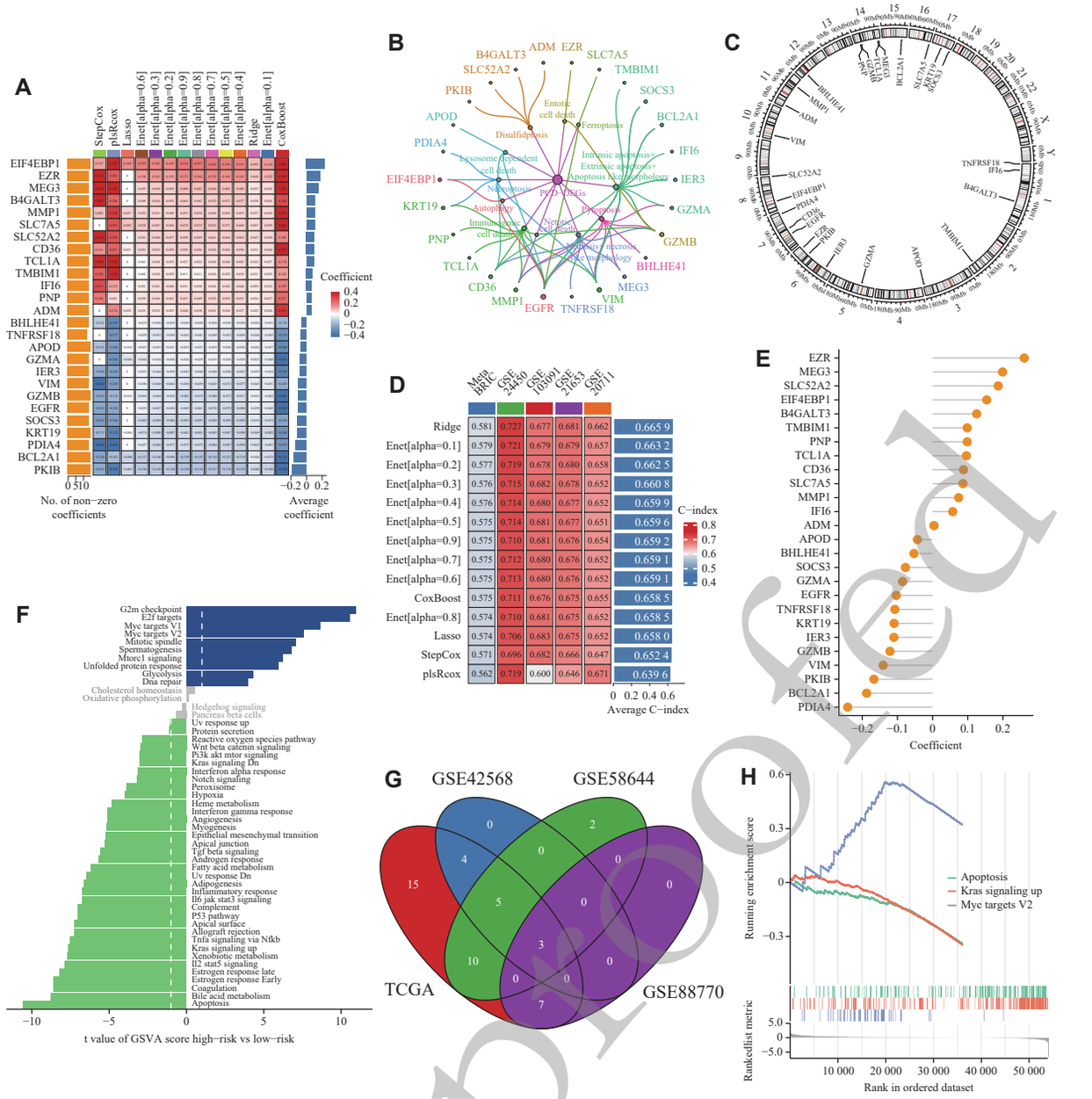


Fig. 4 **Machine learning developed a prognostic risk score.** A: Coefficients of the 26 model genes in 14 models. B: Regulatory network between the prognostic genes and various PCD modalities. C: Circos plot depicting the locations and expression levels of 26 model genes. D: C-index values of the 14 prognostic models across METABRIC, GSE24450, GSE103091, GSE21653, and GSE20711. E: Lollipop chart showing the coefficients of the 26 genes in Ridge model. F: GSVA of the high-risk group and low-risk group categorized by Ridge model. G: Venn plot of enrichment biological functions shared among the TCGA cohort, GSE58644 cohort, GSE42568, and GSE88770 cohort. H: GSEA plots of three hallmark pathways significantly associated with the risk score.

$(0.074 \times MMP1^{exp}) + (0.087 \times SLC7A5^{exp}) + (0.187 \times SLC52A2^{exp}) + (0.088 \times CD36^{exp}) + (0.096 \times TCL1A^{exp}) + (0.098 \times TMBIM1^{exp}) + (0.057 \times IFI6^{exp}) + (0.098 \times PNP^{exp}) + (0.004 \times ADM^{exp}) + (-0.053 \times BHLHE41^{exp}) + (-0.107 \times TNFRSF18^{exp}) + (-0.043 \times APOD^{exp}) + (-0.085 \times GZMA^{exp}) + (-0.110 \times IER3^{exp}) + (-0.141 \times VIM^{exp}) + (-0.121 \times GZMB^{exp}) + (-0.103 \times EGFR^{exp}) + (-0.077 \times SOCS3^{exp}) + (-0.110 \times KRT19^{exp}) + (-0.241 \times PDIA4^{exp}) + (-0.188 \times BCL2A1^{exp}) + (-0.167 \times PKIB^{exp})$. The coefficient of each gene is shown in ***Fig. 4E***. Patients with breast cancer from the TCGA and three independent GEO cohorts were categorized into high-risk and low-risk groups using the median PCD risk score. We used GSVA to explore the biological mechanisms driving various subgroups (***Supplementary Table 5***). The enriched biological processes in the TCGA-BRCA are shown in ***Fig. 4F***, while ***Fig. 4G*** and ***4H*** highlight three shared processes detected in all these datasets. Specifically, the high-risk group showed significant

activation of tumor cell cycle-related pathways and epithelial–mesenchymal transition, indicating enhanced proliferative and invasive capacities. Concurrently, immune-related pathways were suppressed in this group. These results suggest that the high-risk score corresponds to a dual mechanism involving intrinsic tumor malignancy and an immunosuppressive microenvironment. We then analyzed the expression differences of model genes between breast cancer and normal tissues for significance assessment using the Wilcoxon test (***Supplementary Fig. 3A***). Moreover, Kaplan-Meier curves were used to evaluate their impact on overall survival (OS) in breast cancer patients (***Supplementary Fig. 3B***, ***Supplementary Table 6***). Interestingly, all model genes except *MEG3*, *CD36*, and *IFI6* demonstrated significant effects on OS time.

Validation of the clinical predictive performance of the PCD risk score

Breast cancer cases from the TCGA, GSE58644, GSE42568, and GSE88770 databases were divided into PCD high-risk and low-risk groups based on the calculated risk score values. A significant association was observed between the PCD high-risk group and adverse clinical outcomes (***Fig. 5A–5D***). Consistent findings were also obtained in the validation datasets (***Fig. 5E–5I***).

Relationship between the PCD risk score and clinicopathologic features in breast cancer cases

We compared the risk score of each patient with several common prognostic signatures by calculating their C-indices. Notably, the PCD risk score exhibited a significantly higher C-index than other prognostic signatures, such as clinical stage, TNM stage, and grade. Additionally, a comparative analysis of survival status in the training set revealed significant associations between low- and high-risk score groups (***Supplementary Fig. 4A***). Consistent findings were also observed in the validation datasets (***Supplementary Fig. 4B***). Moreover, we analyzed the distribution of risk scores among different clinicopathologic subtypes, as shown in ***Supplementary Fig. 4C–4F***. We benchmarked the prognostic accuracy of the PCD risk score against the established PAM50 (ROR-S) signature. Across the training and validation cohorts, the PCD risk score consistently exhibited superior discriminative ability, yielding higher C-indices compared with the PAM50 signature (***Supplementary Fig. 5***). These results indicate that the PCD-based signature provides more robust survival predictions than traditional molecular classification. Collectively, these results highlight the superior performance of the PCD risk score prognostic model in clinical identification and prediction.

Development and evaluation of nomogram survival models

We further evaluated the independent prognostic significance of the PCD risk score using both univariate and multivariate Cox regression analyses. Subsequently, we identified the PCD risk score as a significant risk factor in the univariate Cox regression analyses (hazard ratio [HR] = 3.10, 95% confidence interval [CI]: 2.41–3.99, $P < 0.001$) (***Fig. 6A***). In multivariate analysis, the PCD risk score as a continuous variable maintained its standalone prognostic value in breast cancer cases, despite adjusting for other covariates (HR = 2.94, 95% CI: 2.26–3.82, $P < 0.001$) (***Fig. 6B***).

Based on the outcomes of the multivariate Cox regression analysis, we constructed a prognostic nomogram model to assess the 1-, 3-, 5-, and 8-year OS using the TCGA dataset (***Fig. 6C***). The calibration curve demonstrated the ability of the nomogram to accurately predict survival rates (***Fig. 6D***). Similar results were observed across all four validation datasets (***Supplementary Fig. 6***). DCA verified that the nomogram significantly outperformed other predictors (***Fig. 6E***). Survival analysis based on the nomogram scores revealed a significant difference between the high- and low-risk groups. To assess the model's performance, we evaluated its prognostic predictive ability, with ROC curves showing a high area under the curve for forecasting 1-, 3-, 5-, and 8-year survival in breast cancer cases (***Fig. 6F***). Consistent outcomes were observed in subsequent validation datasets.

Additionally, we examined each breast cancer subtype individually and created prognostic nomograms for the Basal, LumA, LumB, and HER2 subtypes. The performance of these models in the TCGA cohort was confirmed by DCA and ROC curves in ***Supplementary Fig. 7***. To further validate the risk score in a larger population, we extended the subgroup analyses to the METABRIC cohort. Although the predictive performance varied across specific subtypes compared with the TCGA cohort, consistent prognostic trends were observed in ***Supplementary Fig. 8***. These results support the potential clinical applicability of the model across distinct molecular contexts. The alluvial plot also demonstrated that patients with high risk scores typically had higher nomogram model scores (***Fig. 6G***).

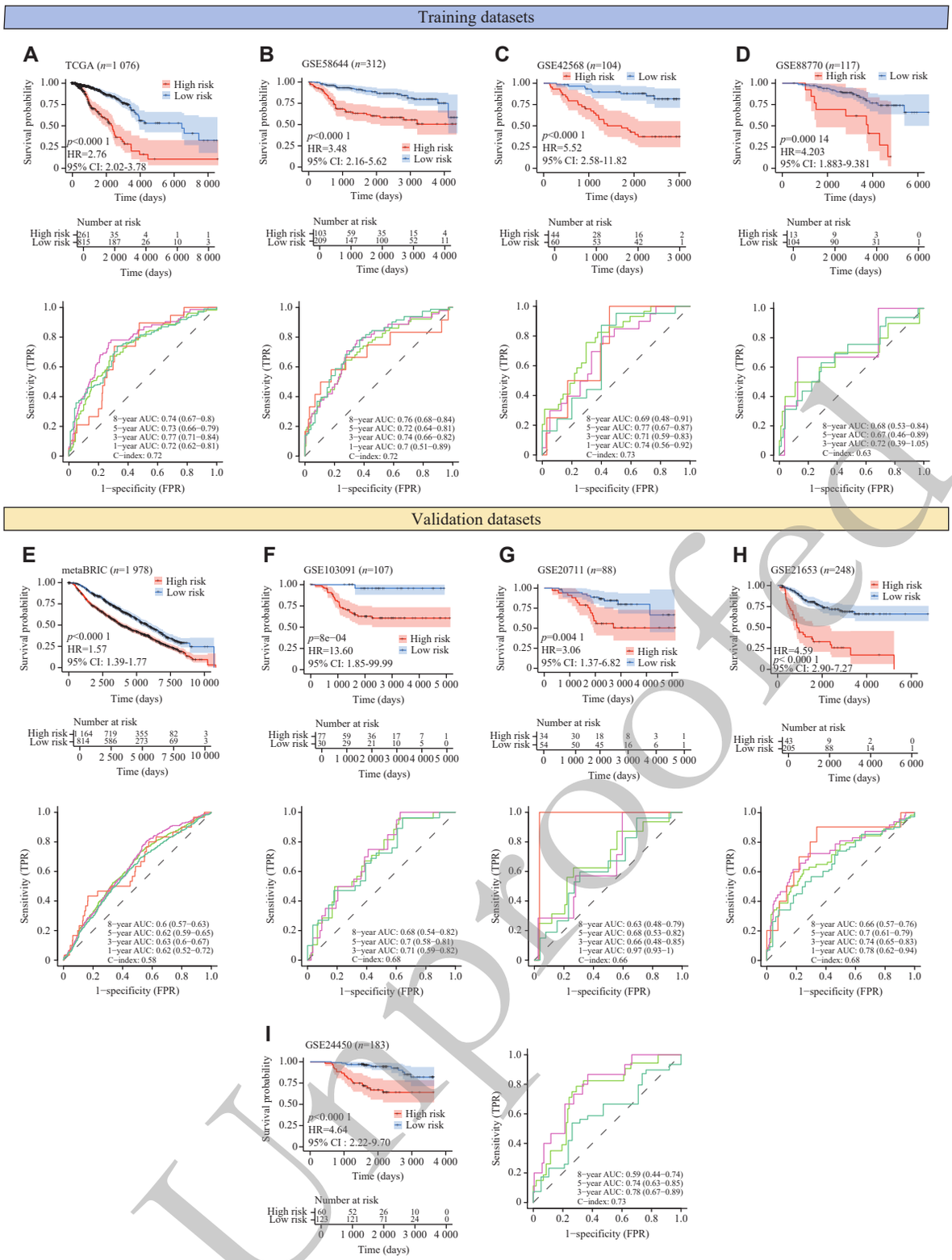


Fig. 5 **Internal training and external validation of the prognostic model.** A–D: Kaplan–Meier survival curves for the high- and low-risk score groups and their corresponding ROC curves among training datasets including TCGA (A), GSE58644 (B), GSE42568 (C), and GSE88770 (D). E–I: Kaplan–Meier survival curves for the high- and low-risk score groups and their corresponding ROC curves among validation datasets including METABRIC (E), GSE103091 (F), GSE20711 (G), GSE21653 (H), and GSE24450 (I).

Single-cell sequencing analysis of PCD model genes

Next, we investigated the expression of model genes at the single-cell level using scRNA-seq analysis. After quality control, cells were retained for subsequent analysis (***Supplementary Fig. 9***). We

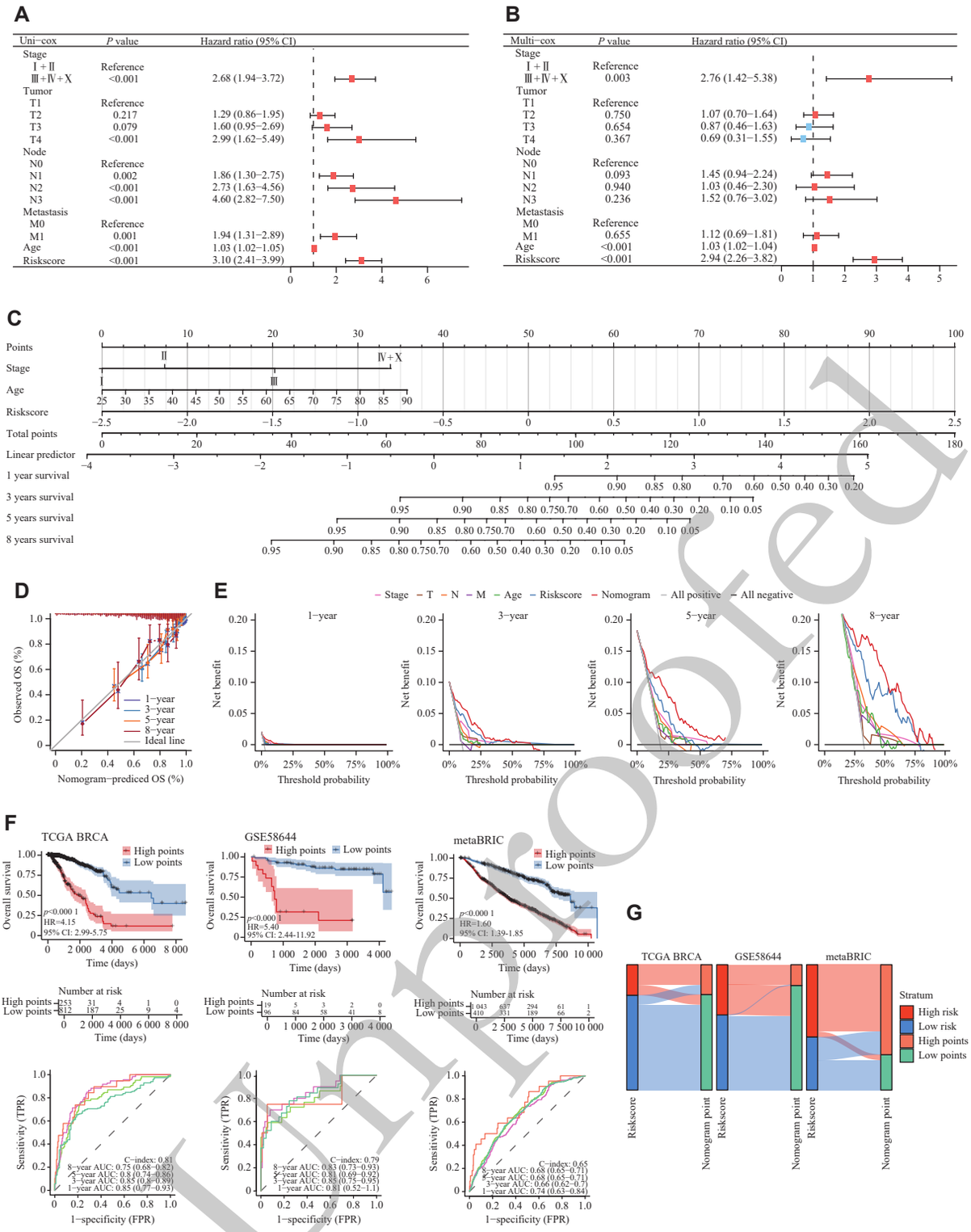
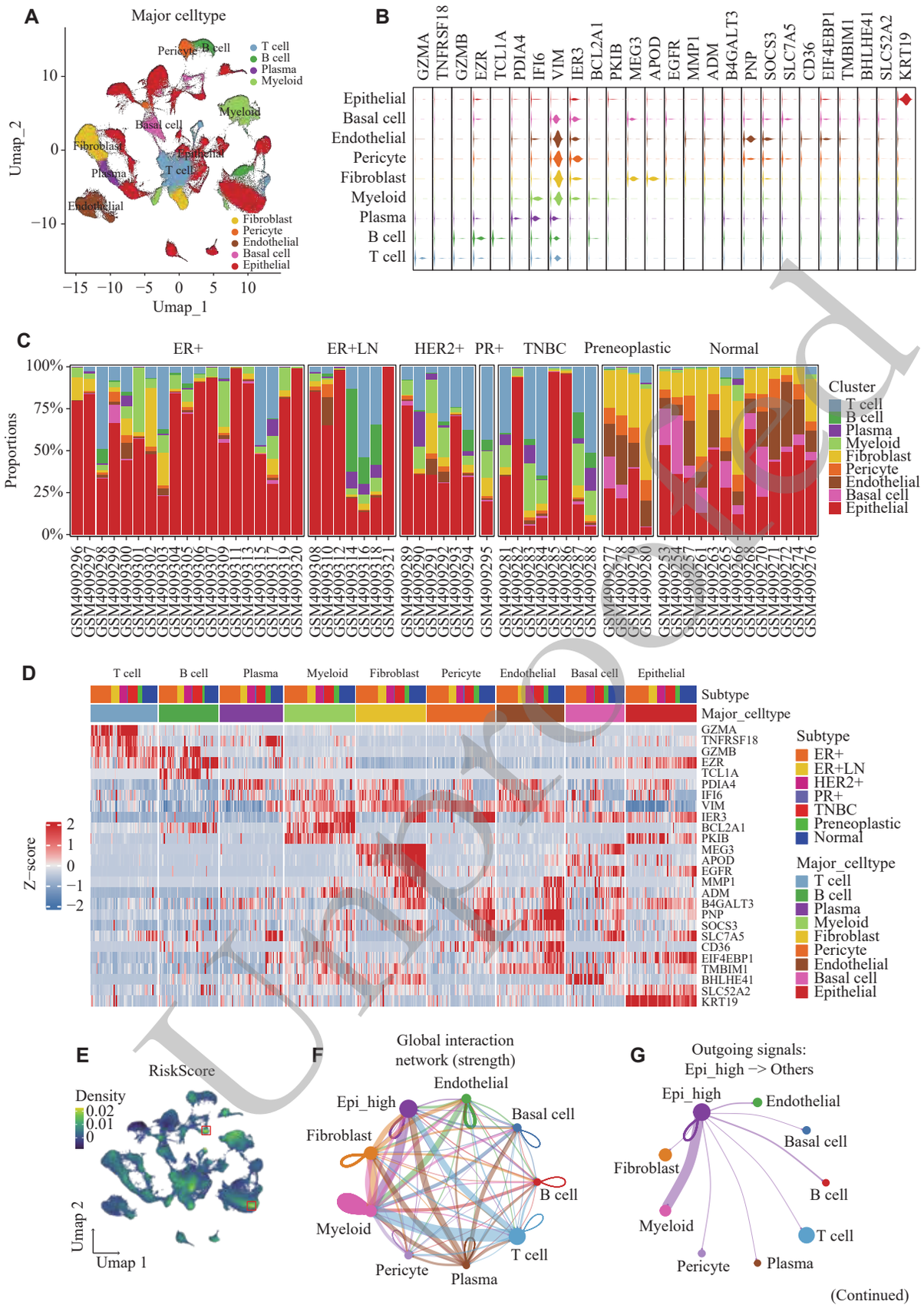


Fig. 6 **Establishment and assessment of the nomogram survival model.** A: Univariate analysis of the clinicopathologic characteristics and risk score in the TCGA-breast cancer cohort. B: Multivariate analysis of the clinicopathologic characteristics and risk score in the TCGA-breast cancer cohort. C: Nomogram established to predict the prognosis of TCGA-breast cancer patients. D: Calibration plots showing the probability of 1-, 3-, 5-, and 8-year overall survival in the TCGA-breast cancer cohort. E: Decision curve analysis (DCA) of the nomogram predicting 1-, 3-, 5-, and 8-year overall survival. F: Survival curves of the two breast cancer groups based on the nomogram score and their corresponding ROC curves in TCGA-breast cancer, GSE58644, and METABRIC. G: Alluvial diagram showing the interrelationship between the risk groups and the groups based on nomogram score in samples from TCGA-breast cancer, GSE58644, and METABRIC.

annotated nine cell subtypes, as illustrated in ***Fig. 7A***. The expression of the 26 model genes within individual cells is shown in ***Fig. 7B***. Next, we explored the expression of the 26 model genes across different clinical samples (***Fig. 7C***). We found that all model genes were significantly differentially



(Continued)

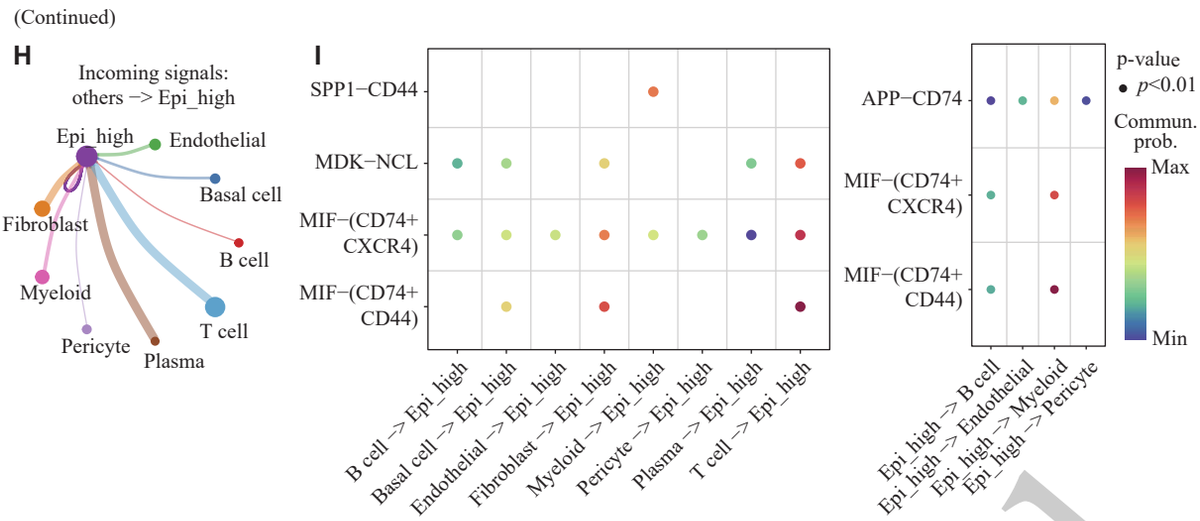


Fig. 7 **Single-cell sequencing analysis of PCD model genes.** A: UMAP plot of nine cell subtypes in the breast cancer dataset. B: Expression levels of the 26 model genes across the nine cell subtypes. C: Proportions of the nine cell subtypes in breast cancer and normal samples. D: Heatmap showing expression levels of the 26 model genes across the nine cell subtypes and different clinical samples. E: UMAP plot showing risk score distribution across cell types. F: Network diagram visualizing overall interaction strengths among nine major cell types. G: Interactions when high-risk epithelial cells (Epi_High) act as ligand cells. H: Interactions when high-risk epithelial cells (Epi_High) act as receptor cells. I: Bubble plot showing enriched ligand–receptor pairs between high-risk epithelial cells and T cells. High-risk epithelial cells as receptor cells (left) and high-risk epithelial cells as ligand cells (right).

expressed in one or more cell types. Additionally, epithelial cells were significantly enriched in ER$^+$ samples compared with other cell subtypes, and fibroblasts were significantly enriched in normal and pre-tumor tissues compared with the other five pathological tissues (***Fig. 7D***).

We further evaluated the PCD risk score at single-cell resolution and calculated a PCD-related risk score for each cell and visualized its distribution on the Uniform Manifold Approximation and Projection (UMAP) embedding (***Fig. 7E***). High risk scores were predominantly enriched in epithelial cells, while immune cells generally showed lower scores, consistent with the bulk RNA association between high score and poor prognosis. Cell-cell communication analysis revealed more active interactions between high-risk epithelial cell subpopulations and multiple immune subsets (***Fig. 7F–7H***). Moreover, several immune-related pairs with strengthened interaction signals were identified in high-risk epithelial-immune crosstalk using ligand–receptor analysis (***Fig. 7I***), suggesting a potential role of high-risk epithelial cells in modulating the immune microenvironment signaling.

The PCD risk score is related to immune function in breast cancer patients

Next, we comprehensively investigated immune cell infiltration patterns across various risk-scored groups in TCGA-BRCA samples using multiple algorithms. Notably, we found a significant negative correlation between the PCD risk score and cells associated with anti-cancer immunity, such as CD8$^+$ T cells and myeloid dendritic cells. Conversely, the PCD risk score displayed a significant positive correlation with M0 macrophages (***Supplementary Fig. 10A***). Additionally, our findings indicated that both the immune score and stromal score were higher in the low-risk score group than in the high-risk score group, as illustrated in ***Supplementary Fig. 10B***. Furthermore, examination of immune checkpoint molecules indicated that all immune checkpoint molecules were downregulated in the high-risk score group, as shown in ***Supplementary Fig. 10C*** and ***D***.

The PCD risk score is associated with therapy response in breast cancer cases

Using the Genomics of Drug Sensitivity in Cancer (GDSC) database, we obtained the half-maximal inhibitory concentration (IC_{50}) values of various drugs in breast cancer samples and investigated the potential correlation between the PCD risk score and drug sensitivity. ***Supplementary Fig. 11A*** illustrates the relationship and significance between IC_{50} values and the prognostic risk score of PCD. Notably, we observed that the high-risk score group exhibited higher IC_{50} values for all drug types compared with the low-risk score group. These included FDA-approved breast cancer drugs, drugs in clinical trials, and those used in basic research (***Supplementary Fig.***

11B). Furthermore, we examined the correlations between model genes, drugs, classical therapeutic targets, and signaling pathways, as illustrated in ***Supplementary Fig. 11C***. The genes *VIM*, *TNFRSF18*, *EGFR*, and *BCL2A1* exhibited a negative correlation with the IC_{50} values of several drugs, while the remaining model genes and the PCD risk score showed a positive correlation with IC_{50} values. These results indicate a possible association between our model genes and drug sensitivity, offering different perspectives on the development of personalized treatment strategies for breast cancer.

Verification of the expression and prognostic significance of hub genes

Subsequently, the differential protein levels of the 26 PCD model genes between breast cancer and normal tissues from clinical samples were validated based on the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset. The results showed that the protein levels of *EZR*, *SLC52A2*, *KRT19*, *PDIA4*, and *PKIB* were overexpressed in breast cancer tissues, while *APOD*, *CD36*, *EGFR*, *PNP*, and *VIM* were downregulated in breast cancer tissues (***Fig. 8A***, ***Supplementary Fig. 12***). We further assessed the protein expression patterns of these ten genes using immunohistochemistry images from the Human Protein Atlas (HPA) database. The staining results for *EZR*, *SLC52A2*, *KRT19*, *PDIA4*, and *PKIB* consistently showed stronger intensity in tumor tissues than in adjacent normal tissues, while the opposite pattern was observed for *APOD*, *CD36*, *EGFR*, *PNP*, and *VIM*, corroborating the CPTAC findings (***Supplementary Fig. 13***). To examine the transcriptional levels of these genes, we performed RT-qPCR on 50 paired clinical breast cancer and normal tissue samples. Among all candidates, *PDIA4* exhibited the most consistent and significant overexpression in tumor tissues (***Fig. 8B–8K***). We therefore focused on *PDIA4* for subsequent immunohistochemical validation (***Fig. 8L–8N***). IHC analysis of the same 50 paired samples confirmed that *PDIA4* protein levels were markedly elevated in breast cancer tissues compared with matched normal controls (***Fig. 8O***). In addition, the IHC scoring data revealed a significant association between high *PDIA4* expression and adverse clinicopathological features. Strikingly, Kaplan–Meier survival analysis demonstrated that patients with high *PDIA4* levels had a markedly shorter overall survival compared with those with low *PDIA4* expression (log-rank test, $P < 0.001$) (***Fig. 8P***). Furthermore, a significant positive correlation was observed between *PDIA4* IHC scores and advanced clinical stages of breast cancer. The median IHC score progressively increased from Stage I to Stage Ⅳ, suggesting a potential role for *PDIA4* in disease progression (***Fig. 8Q***). Collectively, these clinical data substantiate that elevated *PDIA4* expression is not only a frequent event in breast cancer but is also strongly associated with aggressive tumor behavior and poor patient prognosis. These multi-level validation results firmly support *PDIA4* as a robustly upregulated protein in breast cancer, highlighting its potential role in tumor progression.

Integrated analysis reveals the potential regulatory mechanism of *PDIA4*

To elucidate the mechanism of *PDIA4* in PCD regulation, we performed a multi-omics mechanistic analysis. Bulk transcriptomic analysis identified a negative correlation between *PDIA4* and the survival factor *BCL2* (***Supplementary Fig. 14A***). scRNA-seq analysis confirmed the specific co-localization of the *PDIA4-BCL2* axis within the epithelial subpopulation (***Supplementary Fig. 14B***). Crucially, pseudotime trajectory analysis established a clear causal sequence: *PDIA4* upregulation occurs at the early stage of the trajectory, significantly preceding the downregulation of *BCL2* (***Supplementary Fig. 14C***). To identify the downstream mediators, we quantified the signaling output of PDIA4-high epithelial cells, identifying *MK* and *MIF* as the dominant signaling pathways (***Supplementary Fig. 14D***). Circle and bubble plots further revealed that these cells drive the *MIF-ACKR3* autocrine signaling and *MDK-NCL* or *MIF-CD74* paracrine axes (***Supplementary Fig. 14E*** and ***14F***). These results suggest that *PDIA4* suppresses survival signals such as *BCL2* through a secretome-mediated regulatory network.

In vitro experimental verification of the PCD model gene PDIA4

Since *PDIA4* is a crucial gene in necroptosis, we next examined the effects of *PDIA4* on cellular phenotypes. MCF-7 and MDA-MB-231 cells exhibited particularly high levels of *PDIA4* and were therefore selected for subsequent loss-of-function studies (***Fig. 9A*** and ***9B***). The knockdown efficiency of three siRNAs was confirmed in both cell lines (***Fig. 9C*** and ***9D***). Knockdown of *PDIA4* substantially decreased the cell proliferation rate of MCF-7 and MDA-MB-231 cells, as evidenced by CCK-8 and EdU incorporation assays (***Fig. 9E***, ***9F***, ***9H***, and ***9I***). These results were further validated in colony formation assays, where knockdown of *PDIA4* significantly reduced the number of colonies in both

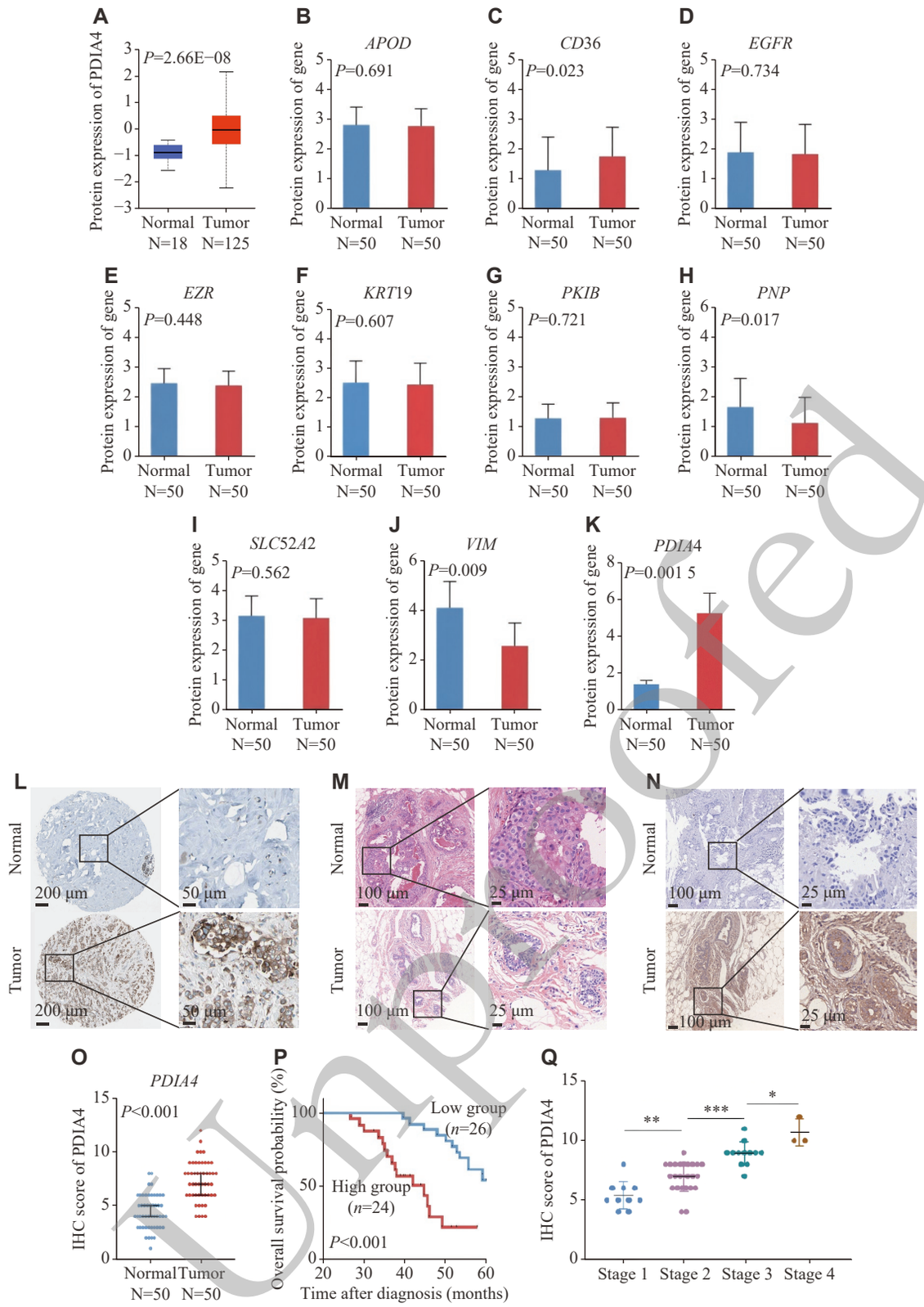


***Fig. 8* PDIA4 is highly expressed and associated with poor prognosis in clinical breast cancer patients.** A: Box plot of *PDIA4* protein expression based on the CPTAC database. B–K: Boxplot of the mRNA expression of PCD model genes using RT-PCR in 50 paired breast cancer and normal tissues. L: Representative immunohistochemistry (IHC) images of *PDIA4* based on the HPA database. Scale bar: 200 μm. M: Representative hematoxylin and eosin (H&E) images of breast cancer tissues based on laboratory samples. Scale bar: 100 μm. N: Representative IHC images of *PDIA4* based on laboratory samples. Scale bar: 100 μm. O: IHC scores of *PDIA4* between breast cancer tissues and paired normal tissues (n = 50). P: Overall survival analysis of *PDIA4* expression in breast cancer patients (n = 50). Q: *PDIA4* expression in patients with different pathological stages (n = 50). Statistical significance was determined by paired Student's t test. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$

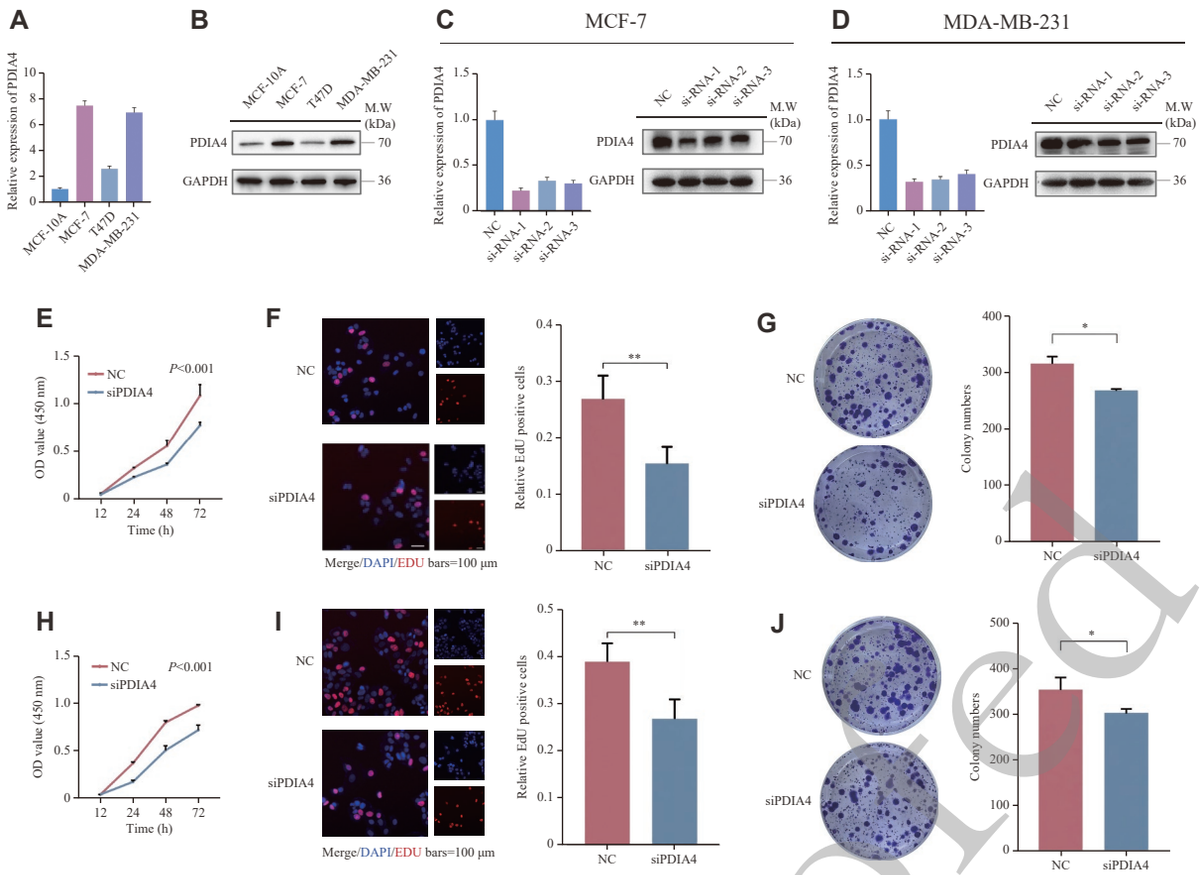


Fig. 9 **Cell proliferation ability of *PDIA4* in breast cancer cells.** A and B: The baseline mRNA and protein expression of *PDIA4* in breast cell lines. C and D: The mRNA and protein expression levels after being transfected with three siRNAs in MCF-7 (C) and MDA-MB-231 (D) cells. E–J: Cell proliferation assessed by CCK-8, EdU, and colony-formation assays in MCF-7 (E–G) and MDA-MB-231 (H–J) cells. Results are presented as mean ± standard deviation. Sample sizes are n = 6 for CCK-8 and EdU, and *n* = 3 for others. Statistical significance was determined by Student's *t* test. $^{*}P < 0.05$, $^{**}P < 0.01$

cell lines (***Fig. 9G*** and ***9J***).

Moreover, Transwell assays showed that knockdown of *PDIA4* dramatically suppressed both the invasion and migration of MCF-7 (***Fig. 10A*** and ***10B***) and MDA-MB-231 (***Fig. 10D*** and ***10E***) cells compared with the NC group. Knockdown of *PDIA4* was also found to increase apoptosis in breast cancer cell lines, consistent with its proposed function in necroptosis regulation (***Fig. 10C*** and ***F***). Altogether, these findings further support that *PDIA4* could function as a potential oncogene involved in breast tumorigenesis.

In vivo experimental verification of the PCD model gene PDIA4

To assess the tumorigenic capacity of *PDIA4* knockdown *in vivo*, we subcutaneously inoculated MCF-7 cells stably expressing either a non-targeting control (NC) or *PDIA4*-specific shRNA into the axillary region of nude mice. Tumors derived from the *PDIA4*-knockdown group exhibited a significantly slower growth rate and formed significantly smaller masses compared with the NC group upon harvest 30 days post-inoculation (***Fig. 11A*** and ***B***). Consistent with the growth curve, the final tumor weight in the *PDIA4*-knockdown group was substantially reduced (***Fig. 11C***). These collective findings indicate that knockdown of *PDIA4* effectively suppresses tumor growth and proliferation *in vivo*, which is potentially associated with the reactivation of the necroptotic cell death pathway.

Discussion

Our study developed a comprehensive machine learning framework incorporating 14 different methods to generate a prognostic PCD riskscore to facilitate prognostic prediction in breast cancer patients. Furthermore, our findings revealed that breast cancer patients with lower PCD risk scores have greater abundances of immune-activated cells, elevated immune scores, and upregulation of immune checkpoints. Breast cancer patients with lower PCD risk scores may also exhibit better responses to

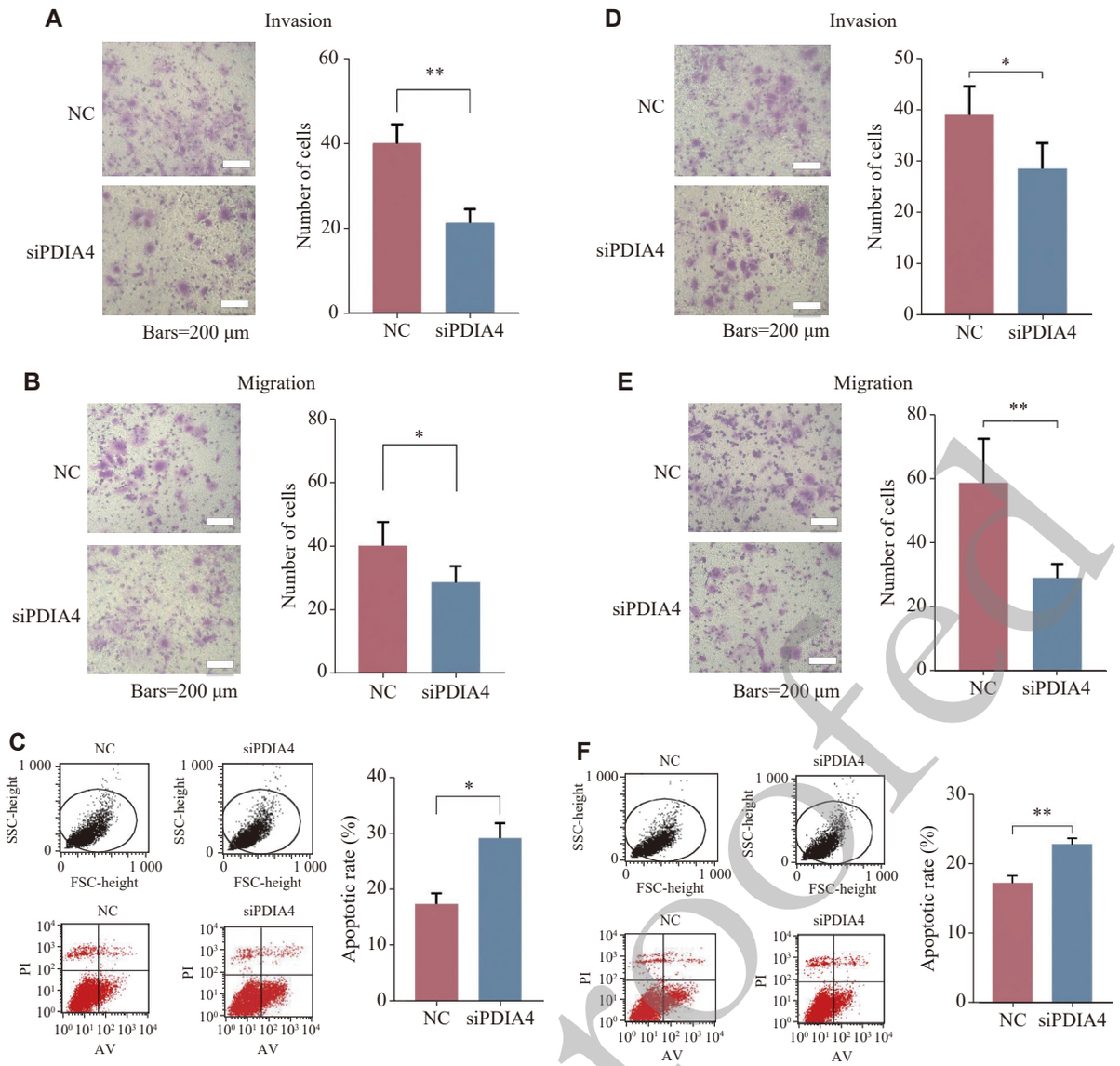


Fig. 10 **Metastatic and apoptosis ability of *PDIA4* on breast cancer cells.** A, B, D, and E: Cell migration and invasion assessed by Transwell assays in MCF-7 (A, B) and MDA-MB-231 (D, E) cell lines. C and F: Apoptosis of MCF-7 (C) and MDA-MB-231 (F) cells assessed using flow cytometry. Error bars indicate the standard error from three independent experiments. Results are obtained from three independent experiments (n = 3). Statistical significance was determined by Student's *t* test. $^{*}P<0.05$, $^{**}P<0.01$

targeted therapies. These findings demonstrate the possible clinical applications of the PCD risk score in directing personalized treatment decisions for breast cancer patients.

The prognostic risk score of PCD was developed based on 26 model genes, including *EIF4EBP1*, *EZR*, *MEG3*, *B4GALT3*, *MMP1*, *SLC7A5*, *SLC52A2*, *CD36*, *TCL1A*, *TMBIM1*, *IFI6*, *PNP*, *ADM*, *BHLHE41*, *TNFRSF18*, *APOD*, *GZMA*, *IER3*, *VIM*, *GZMB*, *EGFR*, *SOCS3*, *KRT19*, *PDIA4*, *BCL2A1,* and *PKIB*. This PCD risk score demonstrates robust and reliable predictive performance for clinical outcomes in breast cancer and functions as an independent prognostic factor. *CD36* has been reported to mediate the import of fatty acids from adipocytes into cancer cells and to activate signaling pathways that promote tumor progression[25]. Knockdown of *IFI6* greatly hindered migration and metastasis with minimal impact on cell proliferation and tumor growth[26]. Consistent with previous research, the current study confirms that these genes serve as positive contributors to poor prognosis within the PCD risk score model. Overexpression of *SOCS3* exhibits an anti-proliferative effect on breast cancer cells, indicating a favorable prognosis[27]. However, a unique subpopulation of immature NK cells with high expression of *SOCS3* was shown to trigger cancer stem cell activation *via* Wnt signaling[28]. Additionally, it was reported that carcinomas form a CXCL12–KRT19 layer to avoid immune assault

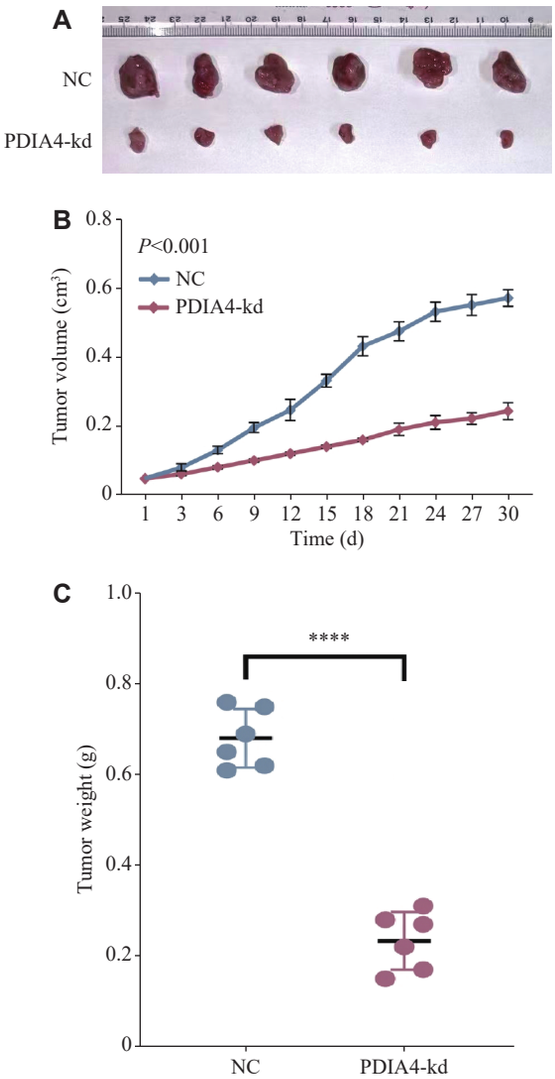


Fig. 11 **Effect of gene *PDIA4 in vivo*.** A: Harvested tumor tissues in the NC and *PDIA4* knockdown groups. B: Tumor volumes in the NC and *PDIA4* knockdown groups. C: Tumor weights in the NC and *PDIA4* knockdown groups. Results are presented as mean ± standard deviation (n = 6 mice per group). Statistical significance was determined by Student's t test. ****$P < 0.0001$.

against cancer[29]. In conclusion, we detected a significant correlation between the expression of model genes and clinical outcomes in breast cancer cases, indicating their viability as prognostic markers for breast cancer. Particularly, disulfidptosis has emerged as a crucial regulator of cancer progression and a promising therapeutic strategy by exploiting redox stress and thiol-disulfide imbalance[30].

Our research uncovered that the immune microenvironment of tumors varies greatly depending on PCD risk score levels. Significantly, high-risk score tumors exhibit fewer infiltrating immune cells compared with low-risk score tumors, including antitumor immune cells as well as immunosuppressive cells. Conversely, M0 macrophages were increased in tumors exhibiting a high risk score. The results demonstrate that immune cell regulation was impaired. Although the low-risk group showed a favorable prognosis, the heightened stromal and immune activities imply latent invasive potential[31]. Assessment of immune checkpoint expression has offered a comprehensive understanding of the diminished immune response in the high-risk score group. We detected decreased levels of suppressive receptors such as CTLA-4, TIGIT, LAG3, HAVCR2, and PD-1, presumably due to T cell exhaustion resulting from chronic immune suppression[32]. Notably, recent evidence suggests that combining low-dose and hypofractionated radiation with anti-PD-1 therapy can reverse such immunosuppression by boosting $CD8^+$ T cell recruitment, thereby effectively preventing lung metastasis in breast cancer[33].

PDIA4 is a critical regulator of PCD. We detected that elevated expression of *PDIA4* in breast cancer tissues was significantly associated with aggressive clinicopathological features and poor patient outcomes. Mechanistically, *PDIA4* may modulate necroptosis sensitivity through the JNK pathway and stabilize the *RIPK1-RIPK3-MLKL* necroptosome complex, amplifying JNK-driven inflammation and promoting radiotherapy resistance[34]. Furthermore, the PLAT-ERK1/2 axis has recently been identified as a pivotal signaling pathway that safeguards cells against multiple PCD modalities, including apoptosis and ferroptosis[35]. Further *in vitro* and *in vivo* experiments demonstrated that reducing *PDIA4* expression in breast cancer cells inhibited their growth and enhanced PCD pathways. These observations imply that *PDIA4* has the potential to be a biomarker for breast cancer prognosis. Notably, *PDIA4* is highly expressed in multiple malignant tumor tissues and correlates with poor prognostic outcomes[36-37]. This suggests the robustness of the PCD risk score within complex biological systems. The innovation of this study lies in elucidating the significant role of *PDIA4* overexpression and cellular PCD pathways in the development and progression of breast cancer.

Several studies have developed various novel models to predict clinical outcomes such as survival in breast cancer. PAM50, which determines intrinsic molecular subtypes and predicts the probability of breast cancer metastasis and recurrence based on the expression of 50 key genes, is now widely accepted internationally[38]. Oncotype DX assesses the expression of 21 genes to predict chemotherapy benefit, thereby guiding treatment decisions[39]. In this study, we identified 26 key genes associated with PCD to establish a prognostic model. Significant

differences in patient survival were observed between the high- and low-risk groups. Furthermore, our model provides insights into the immune microenvironment and sensitivity to therapeutic agents, thereby elucidating the multifaceted role of PCD in the pathogenesis and progression of breast cancer.

This study has several limitations. First, our analyses were predominantly based on retrospective clinical cohorts and computational predictions of tumor immune microenvironment characteristics and drug sensitivity. While these approaches provide valuable insights, the findings require independent validation through experimental methods and prospective cohorts. Second, due to the resources required for comprehensive validation of multi-gene signatures, we focused our experimental validation on the hub gene *PDIA4*. The functional roles of other genes within the signature warrant further investigation. Third, although the molecular subtype-stratified prognostic analyses offer exploratory insights, the reduced sample size within each subgroup may limit statistical power and generalizability. These results should be validated in larger, independent cohorts for each specific subtype.

This study established a machine learning-derived PCD riskscore integrating 26 PCD genes, which outperformed traditional prognostic indicators in predicting breast cancer outcomes. The riskscore-based nomogram enabled clinically actionable risk stratification, which was validated across 9 cohorts. We further functionally identified *PDIA4* as a crucial oncogene driving malignant phenotypes in breast cancer. These findings position the PCD risk score as a dual-purpose tool for prognostication and therapy guidance, with *PDIA4* representing a promising therapeutic target.

Funding

Not applicable.

Author contributions

Jixuan Jiang: Conceptualization, Methodology, Data curation, Visualization, Investigation, Validation, Writing–original draft. **Yiwen Wang**: Methodology, Data curation, Writing–original draft, Visualization. **Yuju Huang**: Methodology, Data curation, Visualization. **Weiwen Yan**: Methodology, Data curation. **Shuwei Li**: Supervision. **Ruoxi Wang**: Supervision.

Acknowledgments

The authors would like to thank all contributors to the study.

Availability of supporting data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040[J]. *Breast*, 2022, 66: 15–23.

[2] Siegel RL, Kratzer TB, Giaquinto AN, et al. Cancer statistics, 2025[J]. *CA Cancer J Clin*, 2025, 75(1): 10–45.

[3] Wang Z, Wang H. Acyl-CoA thioesterase 8 (ACOT8) is a poor prognostic biomarker in breast cancer[J]. *Pharmgenomics Pers Med*, 2024, 17: 403–421.

[4] Zhou Z, Bai N, Zheng W, et al. MultiOmics analysis of metabolic dysregulation and immune features in breast cancer[J]. *Int Immunopharmacol*, 2025, 152: 114376.

[5] Rayson VC, Harris MA, Savas P, et al. The anti-cancer immune response in breast cancer: Current and emerging biomarkers and treatments[J]. *Trends Cancer*, 2024, 10(6): 490–506.

[6] Xiong X, Zheng L, Ding Y, et al. Breast cancer: Pathogenesis and treatments[J]. *Signal Transduct Target Ther*, 2025, 10(1): 49.

[7] Chen Y, Li X, Yang M, et al. Research progress on morphology and mechanism of programmed cell death[J]. *Cell Death Dis*, 2024, 15(5): 327.

[8] Zhang X, Tao T, Qiu Y, et al. Copper-mediated novel cell death pathway in tumor cells and implications for innovative cancer therapies[J]. *Biomed Pharmacother*, 2023, 168: 115730.

[9] Din MAU, Lin Y, Wang N, et al. Ferroptosis and the ubiquitin-proteasome system: Exploring treatment targets in cancer[J]. *Front Pharmacol*, 2024, 15: 1383203.

[10] Karati D, Kumar D. Molecular insight into the apoptotic mechanism of cancer cells: An explicative review[J]. *Curr Mol Pharmacol*, 2024, 17: e18761429273223.

[11] Wang Y, Zhang Z, Sun W, et al. Ferroptosis in colorectal cancer: Potential mechanisms and effective therapeutic targets[J]. *Biomed Pharmacother*, 2022, 153(6): 113524.

[12] Wang S, Wang R, Hu D, et al. Machine learning reveals diverse cell death patterns in lung adenocarcinoma prognosis and therapy[J]. *npj Precis Oncol*, 2024, 8(1): 49.

[13] Cao Q, Zhang JB, Sun DY, et al. Pyroptosis, metabolism, and oxidation in tumorigenesis: Mechanisms and therapeutic implications[J]. *Antioxid Redox Signal*, 2023, 39(7-9): 512–530.

[14] Low HB, Wong ZL, Wu B, et al. DUSP16 promotes cancer

chemoresistance through regulation of mitochondria-mediated cell death[J]. *Nat Commun*, 2021, 12(1): 2284.

[15] Koh M, Lim H, Jin H, et al. ANXA2 (annexin A2) is crucial to ATG7-mediated autophagy, leading to tumor aggressiveness in triple-negative breast cancer cells[J]. *Autophagy*, 2024, 20(3): 659–674.

[16] Xu X, Su Y, Shi J, et al. MicroRNA-17-5p promotes cardiac hypertrophy by targeting Mfn2 to inhibit autophagy[J]. *Cardiovasc Toxicol*, 2021, 21(9): 759–771.

[17] Niu X, You Q, Hou K, et al. Autophagy in cancer development, immune evasion, and drug resistance[J]. *Drug Resist Updat*, 2025, 78: 101170.

[18] Zou Y, Xie J, Zheng S, et al. Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery[J]. *Int J Surg*, 2022, 107: 106936.

[19] Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells[J]. *Nucleic Acids Res*, 2020, 48(W1): W509–W514.

[20] Racle J, de Jonge K, Baumgaertner P, et al. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data[J]. *eLife*, 2017, 6: e26476

[21] Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression[J]. *Genome Biol*, 2016, 17(1): 218.

[22] Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data[J]. *Nat Commun*, 2013, 4: 2612.

[23] Chen B, Khodadoust MS, Liu CL, et al. Profiling tumor infiltrating immune cells with CIBERSORT[M]//von Stechow L. Cancer Systems Biology: Methods and Protocols. New York: Springer, 2018: 243–259.

[24] Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response[J]. *Nat Med*, 2018, 24(10): 1550–1558.

[25] Terry AR, Nogueira V, Rho H, et al. CD36 maintains lipid homeostasis via selective uptake of monounsaturated fatty acids during matrix detachment and tumor progression[J]. *Cell Metab*, 2023, 35(11): 2060–2076. e9.

[26] Roda N, Cossa A, Hillje R, et al. A rare subset of primary tumor cells with concomitant hyperactivation of extracellular matrix remodeling and dsRNA-IFN1 signaling metastasizes in breast cancer[J]. *Cancer Res*, 2023, 83(13): 2155–2170.

[27] Xu J, Shao C, Wang X, et al. circTADA2As suppress breast cancer progression and metastasis via targeting miR-203a-3p/SOCS3 axis[J]. *Cell Death Dis*, 2019, 10(3): 175.

[28] Thacker G, Henry S, Nandi A, et al. Immature natural killer cells promote progression of triple-negative breast cancer[J]. *Sci Transl Med*, 2023, 15(686): eabl4414.

[29] Wang Z, Moresco P, Yan R, et al. Carcinomas assemble a filamentous CXCL12-keratin-19 coating that suppresses T cell-mediated immune attack[J]. *Proc Natl Acad Sci U S A*, 2022, 119(4): e2119463119.

[30] Deng X, Zhang C, Xie J, et al. Disulfidptosis in cancer: From redox stress to therapeutic strategy[J]. *Cancer Gene Ther*, 2026, 33(1): 22–25.

[31] Kather JN, Halama N, Jaeger D. Genomics and emerging biomarkers for immunotherapy of colorectal cancer[J]. *Semin Cancer Biol*, 2018, 52(Pt 2): 189–197.

[32] Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway[J]. *Nat Rev Immunol*, 2018, 18(3): 153–167.

[33] Chen S, Deng X, He X, et al. Preventive effects of low-dose radiation and hypofractionated radiation plus anti-programmed cell death protein 1 on lung metastasis in breast cancer[J]. *Oncol Res*, 2025, 33(3): 687–694.

[34] Tao J, Xue C, Cao M, et al. Protein disulfide isomerase family member 4 promotes triple-negative breast cancer tumorigenesis and radiotherapy resistance through JNK pathway[J]. *Breast Cancer Res*, 2024, 26(1): 1.

[35] He X, Zhang H, Wang Y, et al. Plat safeguards maternally aged oocytes against programmed cell death through activating the Erk1/2 pathway[J]. *Sci China Life Sci*, 2026, 69(2): 506–522.

[36] Shen Y, Zhang N, Chai J, et al. CircPDIA4 induces gastric cancer progression by promoting ERK1/2 activation and enhancing biogenesis of oncogenic circRNAs[J]. *Cancer Res*, 2023, 83(4): 538–552.

[37] Xing F, Song Z, Cheng Z. High expression of PDIA4 promotes malignant cell behavior and predicts reduced survival in cervical cancer[J]. *Oncol Rep*, 2022, 48(4): 184.

[38] Schettini F, Chic N, Brasó-Maristany F, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer[J]. *npj Breast Cancer*, 2021, 7(1): 1.

[39] Syed YY. Oncotype DX breast recurrence score®: a review of its use in early-stage breast cancer[J]. *Mol Diagn Ther*, 2020, 24(5): 621–632.